

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-232908

(43)Date of publication of application : 10.09.1993

(51)Int. Cl.

G09B 5/00

G06F 15/62

G06F 15/62

G06F 15/70

(21)Application number : 04-031804

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 19.02.1992

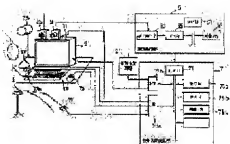
(72)Inventor : KUNO YOSHINORI

(54) INSTRUCTION INPUT DEVICE

(57)Abstract:

PURPOSE: To provide an instruction input device capable of inputting information to a computer or the like based upon the appearance of an operator.

CONSTITUTION: The instruction input device is constituted of a video camera 3 for picking up the images of an operator 1, image memories 57, 75c for storing the images of the operator inputted from the camera 3, a discriminating means for inputting an image relating to operator's operation executed at optional time which is stored in the memories 57, 75c and an image relating to operator's operation executed at time different from the optional time and discriminating the operator's operation by comparing these images, a judging means for judging operator's gesture from operator's operation discriminated by the discriminating means, an input means for specifying the start of the gesture by an input based upon the prescribed operation of the operator 1, and a processing means for giving the concrete expression of the gesture discriminated by the discriminating means when the start of the gesture is specified by the input means.



LEGAL STATUS

[Date of request for examination]

17.02.1999

[Date of sending the examiner's decision of rejection]

08.10.2002

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(51)Int.Cl. ⁵	識別記号	片内整理番号	F I	技術表示箇所
G 0 9 G 5/00	A	8121-5G		
G 0 6 F 15/62	3 8 0	9287-5L		
	4 5 0	9071-5L		
15/70	4 5 0	9071-5L		

審査請求 未請求 請求項の数 4 (全 16 頁)

(21)出願番号 特願平4-31804

(22)出願日 平成4年(1992)2月19日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 久野 義徳

神奈川県川崎市幸区小向東芝町 1 株式会社
東芝総合研究所内

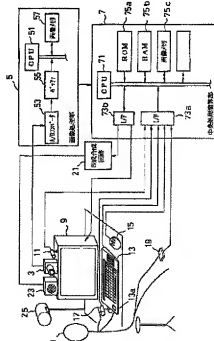
(74)代理人 弁理士 三好 秀和 (外 4 名)

(54)【発明の名称】 指示入力装置

(57)【要約】

【目的】 本発明は、操作者の外観から計算機等に情報を入力することを可能とする指示入力装置を提供することを目的とする。

【構成】 本発明の指示入力装置は、操作者1を撮像するビデオカメラ3と、このビデオカメラ3から入力される操作者の画像を記憶する画像メモリ57、75cと、この画像メモリ57、75cに記憶されるある任意の時刻における操作者1の動作に係る画像と該任意の時刻とは異なる時刻における操作者1の動作に係る画像とを取り込み、これらの画像の比較から当該操作者1の動作を識別する識別手段と、この識別手段で識別された操作者1の動作から当該操作者1の意思表示を判別する判別手段と、操作者1の所定の動作による入力で意思表示の開始が指示されたときには前記判別手段で判別される意思表示を具現化する処理手段とを備えて構成される。



【特許請求の範囲】

【請求項1】 操作者の意思表示を入力し得る第1の入力手段と、

該操作者の所定の動作に関して入力を行う第2の入力手段と、

この第2の入力手段への該操作者の所定の動作に係る入力で前記第1の入力手段による操作者の意思表示を具現化する処理手段とを有することを特徴とする指示入力装置。

【請求項2】 操作者を撮像する撮像手段と、この撮像手段から入力される操作者の画像を記憶する記憶手段と、

この記憶手段に記憶されるある任意の時刻における操作者の動作に係る画像と該任意の時刻とは異なる時刻における操作者の動作に係る画像とを取り込み、これらの画像の比較から当該操作者の動作を識別する識別手段と、

この識別手段で識別された操作者の動作から当該操作者の意思表示を判別する判別手段と、

操作者の所定の動作による入力で意思表示の開始を指示する入力手段と、

この入力手段で意思表示の開始が指示されたときには前記判別手段で判別される意思表示を具現化する処理手段とを有することを特徴とする指示入力装置。

【請求項3】 前記入力手段による入力は前記識別手段によって識別された当該操作者の意思表示開始の指示動作であることを特徴とする請求項1及び2記載の指示入力装置。

【請求項4】 前記処理手段は操作者の所定の動作による入力時刻と同時、若しくは該入力時刻から所定時間経過した時点からの操作者の意思表示を具現化することとを特徴とする請求項1、2及び3記載の指示入力装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は操作者の指示や意図を計算機などの任意の装置に伝達し、操作者が容易に使用できるようにする指示入力装置に関する。

【0002】

【従来の技術】 従来、計算機に各種コマンド等によって指示を入力するにはキーボードやマウスを用いるのが一般的である。さらに多くの情報を伝えたり、あるいはより容易に指示を与えるために、操作者の視線、すなわち操作者がどこを見ているかという情報や、身振り・手振りなどを利用することが考えられている。視線についてはそれを求める専用の装置、例えばアイカメラがあるが、このアイカメラの場合には操作者はアイカメラを着用し、予め整合を取るための調整を行わなければならない、負担を強いられることになる。また、手のサインについてもデータグローブなどの装置を用いて計算機に指示を入力できる。いずれの場合においても、このような

装置を用いる場合は、操作者が装着及び整合を取るための調整に対して負担を強いられしう。

【0003】 そのため、ビデオカメラ等の撮像装置を入力装置に用い、操作者の表情、視線、身振り手振りなどをその画像から認識し、計算機への指示に利用することも考えられている。例えば、顔源、髪部、末永らによるヘッドリダ：画像による頭部動作の実時間検出（電子情報通信学会論文誌、Vol. J74-D-11, No. 3, pp. 398-406）がある。

【0004】 しかし、画像処理による認識は必ずしも確実でないため、例えばある特定の表情で所定の指示を表すと決めておいても、操作者が意図しないときの表情を誤認識して指示であるとして入力が行われる虞がある。また表情の認識そのものに誤りはなくとも、視線や顔の向きでディスプレイ画面上の対象を動かす場合など、操作者がとて、意識することなく顔を動かした場合にも指示と解釈してしまい、操作者の円滑な指示を妨げてしまうことがある。

【0005】 前述したように、画像処理による認識は必ずしも確実ではなく、特に操作者の背景が複雑であったり、照射条件が良くなかったりすると、静止面を入力してから認識処理する従来の方式では対象物の切り出しに失敗することが多くなる。

【0006】 例えば、背景から顔を取り出す場合、顔の部分背景より明るいとして、画像の輝度の大きい所を取り出す方式では、背景に明るいところや暗いところがあったりする場合には良好に動作しない。また、背景の明るさが変わると処理に失敗することが多くなる。

【0007】

【発明が解決しようとする課題】 上述のように、従来の装置では、操作者に負担をかけずに、確実に操作者の外観などを指示入力に用いることはできなかった。

【0008】 この発明はこのような従来の問題点に鑑みて成されたもので、操作者が特別な装置を装着するなどのことなしに、確実に操作者の外観により計算機等に情報を入力することを可能にする装置を提供することを目的とする。

【0009】

【課題を解決するための手段】 上記目的を達成するため本願第1の発明は、操作者の意思表示を入力し得る第1の入力手段と、該操作者の所定の動作に関して入力を行う第2の入力手段と、この第2の入力手段への該操作者の所定の動作に係る入力で前記第1の入力手段による操作者の意思表示を具現化する処理手段とを有することと要旨とする。

【0010】 また、本願第2の発明は、操作者を撮像する撮像手段と、この撮像手段から入力される操作者の画像を記憶する記憶手段と、この記憶手段に記憶されるある任意の時刻における操作者の動作に係る画像と該任意の時刻とは異なる時刻における操作者の動作に係る画

像とを取り込み、これらの画像の比較から当該操作者の動作を識別する識別手段と、この識別手段で識別された操作者の動作から当該操作者の意思表示を判別する判別手段と、操作者の所定の動作による入力で意思表示の開始を指示する入力手段と、この入力手段で意思表示の開始が指示されたときには前記判別手段で判別される意思表示を具現化する処理手段とを有することを有することを要旨とする。

【0011】また、本願第3の発明は、請求項1及び2記載の指示入力装置において、前記入力手段による入力は前記識別手段によって識別された当該操作者の意思表示開始の指示動作であることを要旨とする。

【0012】さらに、本願第4の発明は、請求項1、2及び3記載の指示入力装置において、前記処理手段は操作者の所定の動作による入力時刻と同時、若しくは該入力時刻から所定時間遡った時点からの操作者の意思表示を具現化することを要旨とする。

【0013】

【作用】上述の如く構成すれば、本願第1の発明の指示入力装置は、第2の入力手段で操作者の所定の動作に関して入力が行なわれると、第1の入力手段で入力された操作者の意思表示を具現化するものである。

【0014】本願第2の発明の指示入力装置は、識別手段において、撮像手段から入力され記憶手段に記憶される、ある任意の時刻における操作者の動作に係る画像と該任意の時刻とは異なる時刻における操作者の動作に係る画像との比較から当該操作者の動作が識別される。また判別手段においては、この識別手段で識別された操作者の動作から当該操作者の意思表示が判別される。さらに入力手段で意思表示の開始が指示されたときには、前記判別手段で判別される意思表示が具現化されるものである。

【0015】本願第3の発明の指示入力装置は、入力手段による入力を識別手段によって識別された当該操作者の動作によって行うようにしたものである。

【0016】本願第4の発明の指示入力装置は、操作者の所定の動作による入力時刻と同時に、若しくは該入力時刻から所定時間遡った時点からの操作者の意思表示を具現化するようにしたものである。

【0017】

【実施例】以下、本発明に係る一実施例を図面を参照して説明する。図1は本発明に係る指示入力装置の構成を示したブロック図である。

【0018】図1に示すように、撮像手段としてのビデオカメラ3は操作者1の動作、本実施例においては操作者1の顔の表情の変化を捕らえるものである。

【0019】画像処理部5は、CPU51、A/Dコンバータ53、バッファ55及び画像メモリ57からなり、CPU51は画像処理部5全体のデータの流れ等を制御するものであり、A/Dコンバータ53はビデオ

メラ3で撮像された操作者1の動作に係る画像のA/D変換を行い、入力されるアナログ信号をデジタル信号として出力するものである。このA/Dコンバータ53から出力されたデジタル画像信号はバッファ55を介して画像メモリ57に記憶される。

【0020】中央処理演算部7は、CPU71、入力側インタフェース(I/F)73a、出力側インタフェース(I/F)73b、ROM75a、RAM75b及び画像メモリ75cからなり、CPU71は中央処理演算部7及び当該指示入力装置全体のデータの流れ等を制御するものであり、入力側I/F73aは赤外線センサ11、キーボード13、マウス15、マイクコネクタ17及びフットスイッチ19等と接続されており、これから入力される各種信号の接続を行う。また、出力側I/F73bは、ディスプレイ装置9及び音声合成回路21を介してスピーカ23と接続されており、中央処理演算部7からこれらへの信号の接続を行う。

【0021】ROM75aは中央処理演算部7及び当該指示入力装置全体の制御に係るプログラムを記憶するものであり、また、識別手段、判別手段及び処理手段を含むものである。また、RAM75bは例えば入力側I/F73aを介して入力される赤外線センサ11、キーボード13及びマウス15等から入力される各種信号等、或いはキーボード13によって入力され設定されるプログラム等を入力するものであり、記憶手段としての画像メモリ75cは画像処理部5の画像メモリ57に記憶される画像或いは当該中央処理演算部7で画像処理された画像を記憶するものである。

【0022】ディスプレイ装置9はビデオカメラ3で撮像された操作者1の動作に係る画像や画像処理部5若しくは中央処理演算部7で画像処理された画像等を表示するものである。

【0023】赤外線センサ11は、当該システムの前に操作者1が位置したことを検出するものであり、これによりシステムの動作を自動的に開始せしめることが可能になる。

【0024】キーボード13は、当該システムに各種指示、例えばコマンドの入力或いは注意喚起操作を入力するためのものであり、マウス15、マイクコネクタ17及びフットスイッチ19によっても同様に注意喚起操作等の入力を行うことができる。また、本実施例においては、キーボード13に後述する専用キー13aを設け、使い勝手を良くしている。

【0025】音声合成回路21は、中央処理演算部7で生成され出力される信号にもとづいて所定の音声信号を合成し、スピーカ23から音声を出力するものである。

【0026】スポットライト25は、操作者1の顔と背景との間にコントラストをつけ、画像処理を容易にするもので、操作者1の正面よりは若干斜めの位置から操作者1を照明するように配置される。

【0027】キーボード13、専用キー13a、マウス15、マイクロフォン17及びフットスイッチ19は操作者の所定の動作による入力で意思表示の開始を指示する入力手段及び第2の入力手段を構成するものである。

【0028】次に、本実施例の作用を説明する。なお、操作者が注意喚起操作等の指示をどのような身体的表現を用いて行うかは、いろいろの方法が可能だが、この実施例の説明では、例として顔の特徴の一つである口の形・動きを使用した場合について説明する。

【0029】まず、操作者1がディスプレイ装置9の前に座ると、赤外線センサ11が人間、すなわち操作者1の存在を検知し、中央処理演算部7を起動する。そして、操作者1の方に向けられたビデオカメラ3から画像が、以下で説明する処理中で必要な時点で入力できるように準備される。画像はA/Dコンバータ53によりアナログ/ディジタル変換され、バッファ55を介して画像メモリ57に一旦、蓄えられる。

【0030】この画像が中央処理演算部7のROM75aに格納されるソフトウェアにより処理され、顔・口部分検出、口部分追跡及び口の動き識別等が行われる。このとき、顔・口部分検出、口部分追跡、口の動き識別の3つの処理は、それぞれ時間的に沿ってどう行うかによって、各種の方法が実現できる。以下、順次各処理について説明する。

【0031】まず、図2のフローチャートを参照して顔・口部分検出の処理について説明する。ステップS1で、前述したように前記赤外線センサ11による操作者1の検知により撮像が開始される。この現時点 t_{n+1} の画像とそれ以前の所定の時刻 t_n に撮像された画像をステップS3で取り込み、続いてステップS5でこの取り込んだ画像の両面画につき、差の絶対値を求める。そして、ステップS7で2値化処理を施した後に、この差分結果のうち値の大きい画素を抽出し、これを「1」とする。

【0032】次に、ステップS9で、水平および垂直方向の射影、(この場合水平、垂直の1ラインごとの1の画素数になる)を求める。操作者1の顔は前記2時点(時刻 t_n 、時刻 t_{n+1})の間に多少は動くので、ステップS9における2値化処理により図3に示すような射影部分Sが取り出される。その射影Sは水平方向の射影を示す図、及び垂直方向の射影を示す図のようになる。そこで、射影値の大きいところ、例えば領域Fを取り出すことにより、顔を囲む長方形領域(以下、これを顔領域 E_F という)が取り出せる(ステップS11)。顔の下部については首や肩等の動きで、射影から顔の切り出しにくい場合がある。この場合は顔の縦横の比は概略決まっているので、縦横の値を用いて、顔領域の下限を決定してもよい。

【0033】以上の処理で、射影値の大きい部分が無い場合は、ステップS13で顔検出できなかったとして

ステップS3に戻り、次の時点の画像を入力して処理を繰り返す。

【0034】続いて口領域(図4に示す領域 E_W)の検出を行なう。ステップS15では、顔の位置と、その顔における口の位置は、図4に示すようにある程度の範囲内で位置関係が決まっているので、あらかじめ定められたその位置関係により口領域 E_W を推定し、口の周囲を取り出す(ステップS17)。さらにこの口領域 E_W の中で口の位置をより正確に限定する。これはいくつかの方法が利用できる。

【0035】例えば、一番目の方法は口の周辺部分について顔の切り出しと同様に差分絶対値の画像の射影を求めて、射影値の大きい領域を取り出すやり方である。あるいは、現時点の原画像を使用して、口の周辺領域を2値化して、口の部分(濃度の大きい部分)を取り出すこともできる。これは、2値化結果に連続領域のラベル付けの処理を行い、面積の大きい領域を口として取り出すか、差分画像に対するのと同様に2値化結果に画像の水平・垂直方向の射影を求めて射影値の大きい領域を切り出すことにより実現できる。あるいは、2値化をせずに、原画像の水平・垂直の射影を求め、そこから同様に口の周辺領域を定めることもできる。他にエッジ検出による方法や、スネークス(M. Kass, A. Witkin, D. Terzopoulos: Snakes: Active contour models, Proceedings of 1st International Conference on Computer Vision, pp. 269-276, 1987) による方法を用いて口の輪郭を求めてもよい。

【0036】もし、以上の処理で口に当たる領域が検出できない場合はステップS19でステップS3に戻り、次の時点の画像に対して再度処理するようになる。

【0037】図5は口部分追跡処理のフローチャートである。まず、ステップS21で次時点の時刻 t_{n+1} と前時点の時刻 t_n の画像をそれぞれ取り込む。そして、前時点の時刻 t_n に対して、この画像で口がどこに動いたかを求める(ステップS23)。これには、発明者の他の提案(提案番号13A90X195-1)の中で用いられている追跡法を、前時点の画像で切り出された口の回りの位置に対して適用すればよい。あるいは、前時点の口の回りの画像をテンプレートにして、差分の絶対値の総和や相関係数を、今回の時点の画像の前時点の口の位置の周辺の各点に対して求め、最も類似度の大きい位置を求めてもよく、或いは他の、例えば前述のスネークスを用いて口の動きを追跡してもよい。ステップS25で追跡に成功すれば、同じ処理を繰り返す(ステップS25, 21, 23)。もし、追跡に失敗したら、通常はステップS27に進み、顔・口部分検出処理に戻るようにしておく。

【0038】このような失敗は、操作者が口の形を大きく変えた場合に起こりやすい。特に、追跡に相関係数などの類似度指標を用いた場合や口の形が変化した場合

は、失敗が多くなる。最初の方法やスネークスによる方法では、そのような口の形の変化に対しての失敗は少ない。

【0039】図6に口の動き識別処理のフローチャートを示す。この処理では前後する2時点、例えば前述した時刻 t_0 と時刻 t_{0+1} における画像を用いる。その画像入力の時点の前後関係等については後で述べる。ここでは、2時点の画像が与えられたものとして処理法を説明する。なお、この口の動き識別処理部分は3時点の画像を用いる方法でも同様に実現できる。これについても後で説明する。

【0040】この図6に示す口の動き識別処理により操作者の意思が最終的に解釈されるため、ここでは簡単に確実な処理を用いる。

【0041】まず、口による意思伝達として、通常は口の開閉による動きと、口の左右方向及び上下方向への動きを使用することができる。但し、口の上下左右の動きは、実際には口が動くわけではなく顔全体が動くために、口の位置が変わるものである。

【0042】始めに、ステップS31で追跡処理で求められている口周辺の領域について、2時点、例えば前述した時刻 t_0 と時刻 t_{0+1} における画像間の差分を取る。ここでは、デジタル画像の中で明るい部分は大きな数値、暗い部分は小さい数値を取るものとする（ここで逆にして明るい部分は小さな数値、暗い部分は大きな数値としてもよく、その場合は以下の説明で符号が逆になる）。以下、このステップS31で求められる差分は動作をした後の時点の画像から前の時点の画像を引いたものとして説明する。

【0043】図7は、このようにした場合の口の開閉、口の上下左右の運動の際の差分値のパターンを模式的に示したものである。図7においては、実線が運動前、破線が運動後の形を示す。差分結果の正負の概略を図は示している。従って、処理としてはこれらのパターンを識別すればよい。そのために、差分結果から正と負それぞれ絶対値が一定のしきい値以上の部分を2値化により取り出す（ステップS33）。そして、それぞれにつき、顔の検出で用いたように、水平と垂直方向の射影を求める（ステップS35）。2値化をせずに差分結果を正負にわけ、それぞれについてそのまま、あるいは絶対値が一定値以上のものを選んで射影をとってもよい。射影の結果は、模式的に示すと、口の各様の動きに対して図8乃至図10のようになる。この図でわかるように水平・垂直の射影の結果の中の山の存在の組み合わせと山の幅と位置関係で、口の動きが識別できる。そこで、それぞれの射影結果について山の存在・幅・位置を測る（ステップS37）。そして、図11に示すような判定テーブルで口の動きを識別する（ステップS39）。

【0044】次に、以上の処理を組み合わせて、操作者の注意喚起操作を加えて、本発明を実施する方法について

述べる。図12乃至図18は処理時間的に組み合わせる方法の例を示したものである。これらの図の中で、中間の長さの垂直線分は顔・口部分検出の処理を行う時点を示す。線は1本だが、この線で示した時点と次の時点の画像を用いて処理を行う。短い線分は追跡処理を行っていることを示す。線分で示される時刻の画像とその前の時点の画像を用い、現時点の口の位置を追跡する。三角と四角のしるしは口の動き識別処理で用いる2時点の画像の入力時点を示す。

【0045】長い線分は注意喚起操作の起こった時点を示す。操作者は、自分の意思を伝えたいことを、この操作により示す。これはキーボード13により、ある特定のキーあるいはキーの組み合わせを押すことにより行う。通常のキー（の組み合わせ）とは別に専用キー13aを設けてもよい。他に、マウス15のある定まった押し方で行っても、あるいはフットスイッチ9を付加して、その操作により行ってもよい。注意喚起の操作により、指示の意思を伝えるときにも、その時点の画像を入力し、それと他の時点の画像の間の処理により、背景等の周囲環境の変化に強い処理が実現される。

【0046】注意喚起操作は、上に述べたようにキーボード13を押すなどで行われるが、キーボード13などを押した後、直ぐに離す必要はなく、しばらく押し続けてもよいように入力制御部11で入力を制御してもよい（通常は、キーボード13は押した時点だけ入力される、このような形になっている）。これは、注意喚起操作をしてからある意思表示の動作を操作者がするが、その動作と同時にキーボード13を離す動作をするのは、操作者にとって使いにくい場合があるからである。

【0047】図12は基本的な実施法である。顔・口検出が成功すると追跡処理が始まる。追跡に失敗すると再び顔・口検出が起動される。あるいは検出に失敗しなくても、時々顔・口検出を行い、ある処理を修正してもよい。そして注意喚起操作を行うとその時点（三角印）と次の適当な時間後（四角印）の画像を入力して、口の動き識別処理を行う。この場合操作者は、注意喚起操作を行ってから、意思伝達の動作を行う。

【0048】図13は図12と同様であるが、動きの識別の処理に注意喚起操作の直前（あるいは一定時間前）の時点の追跡に用いた画像と注意喚起操作の時点の画像の2枚の画像を用いる。この場合は操作者は意思表示の動作の完了とともに注意喚起操作を行う。完全に動作が完了してから注意喚起操作を行うという形でもよいが、操作者がある意思表示をしようと思うと同時に注意喚起操作の動作にも入ろうと思い、前者の実際の動作を開始すると、その直後は同時に後者の動作を開始するようにしてもよい。実際には注意喚起操作から画像入力には少しの時間遅れがあり、これを適当に調節して、使いやすいタイミングを設定することができる。この点は、この実施例の場合に限らず、この発明の他の実施例にも適

用することができる。すなわち、注意喚起で入力される画像については、注意喚起操作から適当な時間遅れをもって入力されてよい。

【0049】図14は追跡の部分を省略した例である。この場合、注意喚起操作の時点の画像と、その後のある時点（図中の丸印）で顔・口部分を求め、さらにその後のある時点の画像と、前のいずれかの時点の画像を用いて、口の動作識別を行う。

【0050】さらに簡略化した方法として、顔・口検出と口の動作識別に同一時点の画像を用いるようにしてもよい。この場合は両方の処理とも、注意喚起操作時点の画像とその後のある時点の画像を用いて処理を行うことになる。口の動きでなく、顔全体の動きで意思を伝える場合には、この方法で十分である。顔の動きで意思を伝える方法の例については後で述べる。

【0051】図15も追跡の部分を省略した例である。この例では、画像の入力だけは図中のX印で示すように、たえず適当な間隔で行っておく。そして、直前の適当な枚数だけメモリ57に蓄えおくようにしておく。注意喚起操作があったら、その前の適当な時間の蓄えられた画像（図中の丸印）と注意喚起操作の時点の画像を用いて、顔・口検出を行い、さらに注意喚起操作時点とその次の時点の画像を用いて、口の動き識別を行う。この方法は図14の方法に比べ、適切な処理を行うのに必要な時間間隔を自由に設定できる。この方法と図13の方法の合成したものととして、四角印の入力時点を図中の矢印のように注意喚起操作の前にもってきて、図13の場合のような操作を行う方法もある。また、図14の説明の後半で述べたように、顔・口検出処理を省略する方法もある。この場合は図13の場合のような操作法になる。

【0052】図16は追跡は行わないが、顔・口検出は適当な時間ごとに行うという方法である。図14の場合に比べて、顔・口部分検出処理を注意喚起操作後に行う必要がないので、口の動き識別の処理に直ぐ入れ、早い応答が可能である。

【0053】図17は図15と図16の場合を合わせたもので、追跡は行わないが、顔・口検出の間にも画像入力しておき、四角印と三角印の画像を用いて口の動き識別を行い、図13のような操作をするものである。

【0054】図18は図19のように操作者が口（顔）を上下左右に動かすことによりディスプレイ9上のアイコンを動かしたり、マルチウィンドウのウィンドウの移動や、その他のポインティングを行うような場合に使用される方法である。まず注意喚起操作が起こると、その時点と次の時点の画像で顔・口の位置検出を行い、その後、口の追跡処理が実行され、その動きに応じてディスプレイ画面との操作対象が動く。操作者が満足する位置に移動などができたら、操作者は再び注意喚起操作を行う。この操作の起こった時点の結果が、最終的な移動等

の結果として採用される。注意喚起操作としては毎回キーボード13を押すような操作をしてもよいが、最初の操作でキーボードを押し、それを押し続け、2回目の注意喚起操作は、それを離すことにより行うように入力側インタフェース73aを設定しておいてもよい。また、顔・口検出処理を省略する実施例も可能である。その場合は丸印の画像入力は不要である。

【0055】以上が本発明の基本的な実施例であるが、本発明は以下に述べるような方法をその構成手段の実施法としてもよい。例えば操作者の意思伝達に顔の大きな動きを用いる場合は、図20にフローチャートを示す方法を用いてもよい。まず、ステップS41で注意喚起操作の時点とその後（図17のような方法を用いる場合は前の時点でも可、この場合は図15の場合のような使用法になる）の2枚の画像の差分をとる。以下、図2に示したステップS3乃至ステップS11と同様にして、顔領域を決定する（ステップS41乃至ステップS47）。そしてステップS49で顔領域の縦と横の比を求め、あらかじめ定めておいた値より、横の方に長いなら顔を横に動かした、縦の方に長いなら縦に動かしたと判定する。前に述べたように、顔の縦の領域が求めにくい場合は、横に動いたら横、そうでなければ縦と判定するようにしてもよい。

【0056】顔を動かす方法で、顔を縦に何回か振ってうなずいたり、横に何回か振って否定を表したりするような場合は、2枚の画像ではなくて、何枚かの画像を注意喚起操作後（あるいは前）に入力して、連続時間あるいは他の適当な組み合わせの画像間で差分を2値化した結果を加算し、顔領域を決定してもよい。このようにすれば、このような動作を用いる場合の意思の判定がより確実になる。

【0057】次に口の動き識別の処理の別の実施例について述べる。この方法では、図21のように口の領域の中に、さらに小さな領域を設ける。具体的には口の中央領域ec、上側領域eu、右側領域er、左側領域el、下側領域edを設ける。図22に示すフローチャートにおいて、まず各小領域eについて2時点の画像の差分を求める（ステップS51）。そして、小領域eごとに、小領域eの全画素に対する差分結果の総和、あるいは適当なしきい値より絶対値が大きい正の部分の画素数と負の部分の画素数を求め、両者の差を計算する（ステップS53）。その結果に対して図23の判定表を利用することにより、口の動き識別を行う（ステップS55）。図23では+は領域の和が正で絶対値が一定値以上、あるいは正の画素数が負の画素数より一定値以上多いことを示す。-は逆である。斜線は+や-を付したものの絶対値が小さいものである。

【0058】これまでの実施例では基本的には2枚の画像を用いていたが、以下のように3枚の画像を用いる方法もある。この場合も、画像2枚の場合と同様にいるい

11

る時点で画像を入力する方法が実施できるが、ここでは代表的な例を用いて、顔の動き検出を例に説明する。

【0059】ここでは、図24のタイミングチャートに示すように、画像は抱え入り入力されており（図中の画像入力時点X印）、注意喚起操作が三角印の時点から起こるとする。本実施例における顔の動き検出処理には、注意喚起操作の時点の直前の時間前の画像（図中の丸印）、注意喚起操作の時点の画像（図中の三角印）、それにその後の直前の時間前の画像を用いる（図中の四角印）。

【0060】また、操作者1は注意喚起操作の後、意思表示動作を行うものとする。さらに、この例では意思表示動作は顔を左右に動かす（左右に回す）、上下に動かす（見上げる、うなずく）というような動作とする。

【0061】処理は始めに、三角印と丸印の差分画像を求め（ステップS61）、そこから図20のステップS43乃至ステップS47と同様に顔の領域を求める（ステップS63）。注意喚起操作の前なので、顔はあまり動いていないので、この処理では差分により顔全体の輪郭がでる。次に四角印と三角印の時点の画像の差分を求め（ステップS65）、これも図20のステップS43乃至ステップS47と同様に、顔領域を求める（ステップS67）。そして、顔領域の大きさについてステップS67の結果はステップS63の結果に対して、縦方向と横方向のどちらにより大きくなったかを求める（ステップS69）。大きくなった方向に顔が動かされたかと判定する。この方法では、動作前の大きさで動作後の大きさの比較を用いるので、正確さが向上する。また、この場合は近い時点の比較なので、顔の画像上での位置の絶対値の変化で動きを判定することもできる。顔領域の右端、左端、上端、下端の変化した方向に動いたと判定する。下端は求めにくい場合があるが、そのような場合は上端がどちらに動いたかで判定できる。

【0062】さらに円滑な、いわゆるヒューマンインターフェイスを実現するために本発明を適用することもできる。以下、図26を参照してこのような装置構成を説明する。

【0063】例えば、図26(a)に示す中央処理演算部7で操作者1に対して所定のメッセージをディスプレイ装置11に表示出力し、或いは音声合成装置21で作成された音声スピーカ23を介して出力する等して、質問を行ない操作者1の返答を要求する場合、中央処理演算部7は返答要求後の短時間の後に操作者1から反応があることが期待できる。そこで質問を出力したことを注意喚起操作の時点と見なし、続いて意思表示の認識が可能な状態となり、待機状態に入る。従って、この場合には操作者1は特に注意喚起操作をする必要はない。

【0064】また図26(b)に示すように、中央処理演算部7が特に明確な返答要求を行わない場合などは、注意喚起操作を操作者1が行うようにする。

【0065】このようにすれば、操作者1は聞かれれば

12

そのまま返答し、何か操作者1から特に言いたいとき、すなわち入力を行ないたいときには注意を引いて確実に伝達するという形になり、操作者1における操作感向上する。

【0066】また上記実施例では、注意喚起操作としてキーボード13への入力等による操作を用いたが、マイクロホン17を用いて、操作者の音声或いは拍手、足踏み、手で他のものを叩く等の他の音を検知して、それを注意喚起操作としてもよい。この場合、キーボード13からの入力に比べ音声認識に対する処理時間が掛かることから、意思表示の画像入力開始時点のタイミングには注意を要する。

【0067】また、本実施例では意思表示を検知するセンサとしてビデオカメラ3を用いるようにしたが、任意の他のセンサを用いてもよい。例えば、マイクロホン17を用いて音声や他の音を使ってもよい。この場合、注意喚起操作の時点で、音を入力し、この入力された音声を認識する。このとき音声による入力は画像による場合よりも、認識に係る処理時間を短くすることができる。

また、音声の認識には通常の音声認識手段を用いてもよいが、画像の認識処理の場合と同様に、注意喚起操作時点の音とその後（あるいは前）の音を入力し、その2点（あるいはそれ以上の数の点）での大きさや周波数の差を求め、それを意思表示と対応するようにしておけば、簡単な処理で確実な意思伝達が実現できる。例えば、あるキーを押して注意喚起操作を行った場合は音の大きさが、後の時点の方が大きくなったら画面上の対象物が上に動くようにする。

【0068】さらに上記各実施例では、画像はすべてメモリに記憶してから、計算機のソフトウェアで処理するようにしている。これを、最新時点の画像はメモリに一旦記憶すること無く、入力と同時にハードウェア的に、記憶された前時点の画像と差分を取るようにしてもよい。

【0069】また、実施例では意思伝達処理の部分で操作者が使う中央処理演算部7を用いて実現しているが、この意思伝達処理専用のハードウェアを構成して実現するようにしてもよい。

【0070】以上、具体的に説明したように本実施例によれば、操作者は特別な装置や器具を身体に装着すること無く、また負担を強いられること無く指示を入力することができる。これにより、従来の意思伝達の際の誤りの多発を、本実施例による注意喚起操作により意思伝達に必要な時点を描え、その時点と後（あるいは前）の時点の画像内の変化を使うことにより周囲環境の変化などに影響されず確実に意思伝達を行うことができる。

【0071】尚、上記の実施例では顔及び口の動きを意思表示に適用した場合を例にとって説明したが、本発明はこれに限定されること無く、例えば目、耳、鼻或いは手、指等の適宜の部位を単独または適当に組み合わせ

て意思表示に適用することができる。これらの場合についても、画像間の差分で対象の外形を出してその形を射影などで求めたり、対象の回りに小領域を設定し、その中の明るさ変化を求めることにより、実施例の場合と同様に実現できる。

【0072】

【発明の効果】 上述したように本発明は、操作者の所定の動作を注意喚起として操作者の意思表示を具現化するように構成したので、操作者に対して負担を強いること無くかつ周囲環境の変化などに影響されること無く指示

【図面の簡単な説明】

【図1】 本発明の一実施例の処理の構成を使用形態と共に示すブロック図である。

【図2】 顔・口部分の検出処理の手順を示すフローチャートである。

【図3】 顔の射影から顔領域を検出する様子を説明するための説明図である。

【図4】 口の射影から口領域を検出する様子を説明するための説明図である。

【図5】 口部分追跡処理の手順を示すフローチャートである。

【図6】 口の動き識別処理の手順を示すフローチャートである。

【図7】 口の動きを各パターン毎に説明する図である。

【図8】 図7に示す各パターンの内、横方向の口の動きに伴う射影の形を示す図である。

【図9】 図7に示す各パターンの内、上下方向の口の動きに伴う射影の形を示す図である。

【図10】 図7に示す各パターンの内、開口動作に伴う射影の形を示す図である。

【図11】 図7、図9乃至図10における口の動き識別のための判定表を示す図である。

【図12】 画像入力・処理のタイミングチャートである。

【図13】 画像入力・処理のタイミングチャートである。

【図14】 画像入力・処理のタイミングチャートであ

る。

【図15】 画像入力・処理のタイミングチャートである。

【図16】 画像入力・処理のタイミングチャートである。

【図17】 画像入力・処理のタイミングチャートである。

【図18】 画像入力・処理のタイミングチャートである。

【図19】 図18に示すタイミングチャートの場合の操作の状態を説明する図である。

【図20】 顔の動きによる意思伝達処理の手順を示すフローチャートである。

【図21】 口の回りに設定した小領域の説明図である。

【図22】 小領域を用いた口の動き識別処理の手順を示すフローチャートである。

【図23】 小領域を用いた口の動き識別処理のための判定表を示す図である。

【図24】 3時点の画像を用いる場合の画像入力・処理のタイミングチャートである。

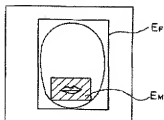
【図25】 3時点の画像を用いる場合の処理の手順を示すフローチャートである。

【図26】 注意喚起操作を使い分ける場合を説明する図である。

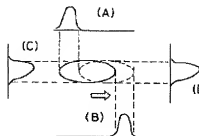
【符号の説明】

- 1 操作者
- 3 ビデオカメラ
- 5 画像処理部
- 7 中央処理演算部
- 9 ディスプレイ装置
- 11 赤外線センサ
- 13 キーボード
- 15 マウス
- 17 マイクロフォン
- 19 フットスイッチ
- 21 音声合成回路
- 23 スピーカ
- 25 スポットライト

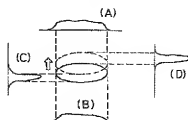
【図4】



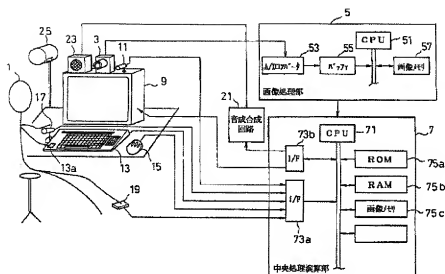
【図8】



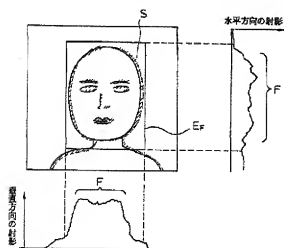
【図9】



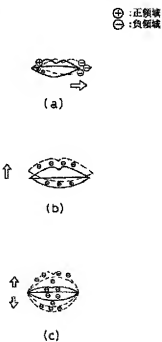
【図1】



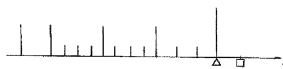
【図3】



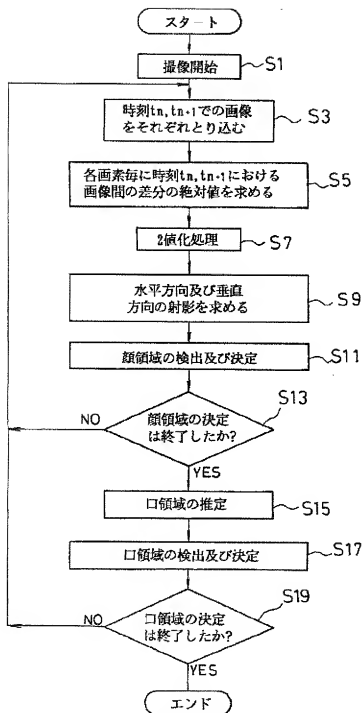
【図7】



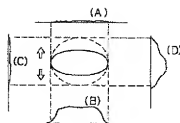
【図12】



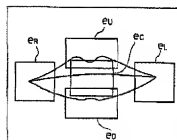
【図2】



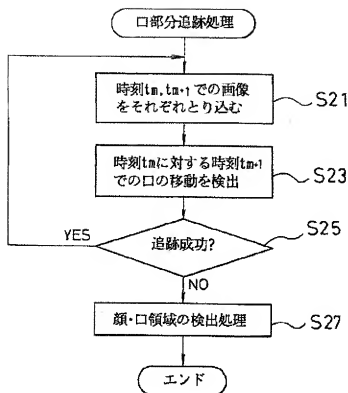
【図10】



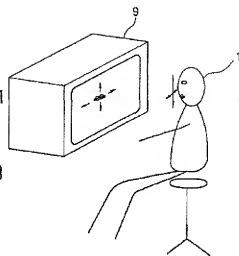
【図21】



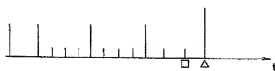
【図5】



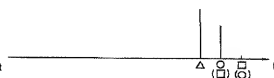
【図19】



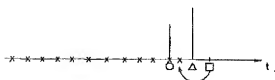
【図13】



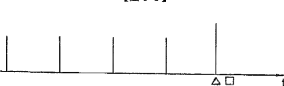
【図14】



【図15】



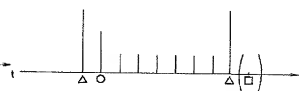
【図16】



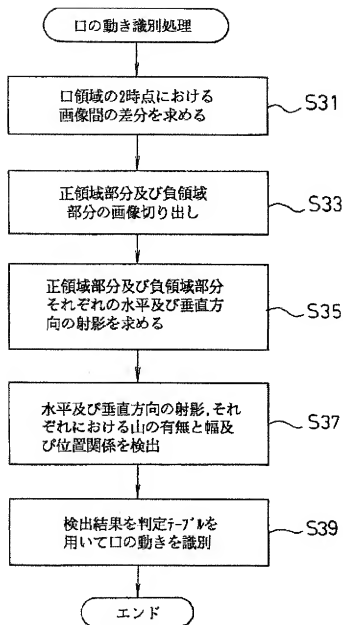
【図17】



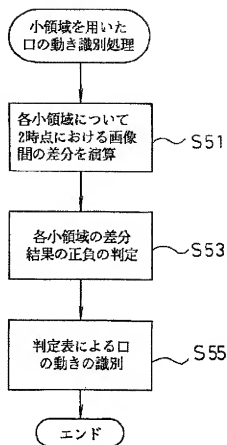
【図18】



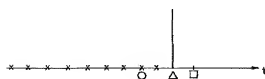
【図6】



【図22】



【図24】

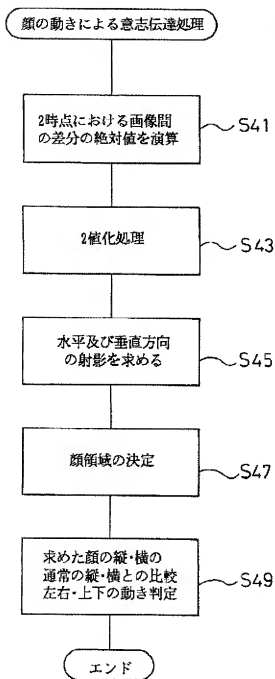


【図11】

動き	垂直射影				水平射影			
	正	負	幅	位置関係	正	負	幅	位置関係
右	○	○	狭	正が負の左	○	○	狭	同位置
左	○	○	狭	正が負の右	○	○	狭	同位置
上	○	○	広	同位置	○	○	広	正が負の下
下	○	○	広	同位置	○	○	広	正が負の上
開	X	○	広		X	○	広	
閉	○	X	広		○	X	広	

○:山あり
 ×:大きな山なし

【図20】

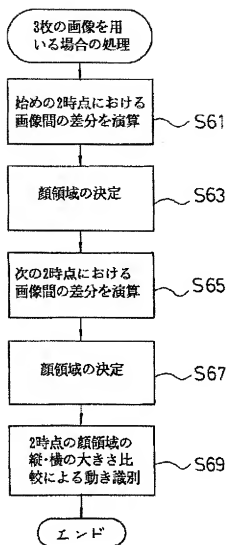


【図23】

動き	小領域	A	B	C	D	E
右		+	/	-	/	/
左		-	/	+	/	/
上		/	-	/	+	/
下		/	+	/	-	/
開		/	/	-	-	-
閉		/	/	+	+	+

+++: 変化する
 ---: 変化する
 / : 大きな変化なし

【図25】



【図26】

